Antigen presenting cells that contain an optimized recombinant genetic vaccine vectors can be identified by, e.g., detecting expression of a marker gene that is included in the vectors.

The invention also provides methods of evolving a bacteriophage-derived vaccine delivery vehicle to obtain a delivery vehicle having enhanced ability to enter a target cell. These methods involve the steps of. (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which encodes an invasin polypeptide, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant invasin nucleic acids; (2) producing a library of recombinant bacteriophage, each of which displays on the bacteriophage surface a fusion polypeptide encoded by a chimeric gene that comprises a recombinant invasin nucleic acid operably linked to a polynucleotide that encodes a display polypeptide; (3) contacting the library of recombinant bacteriophage with a population of target cells; (4) removing unbound phage and phage which is bound to the surface of the target cells; and (5) recovering phage which are present within the target cells, wherein the recovered phage are enriched for phage that have enhanced ability to enter the target cells. Again, if further optimization is desired, the methods can include the further steps of (6) reassembling (&/or subjecting to one or more directed evolution methods described herein) a nucleic acid which comprises at least one recombinant invasin nucleic acid obtained from a bacteriophage which is recovered from a target cell with a further pool of polynucleotides to produce a further library of recombinant invasin polynucleotides; (7) producing a further library of recombinant bacteriophage, each of which displays on the bacteriophage surface a fusion polypeptide encoded by a chimeric gene that comprises a recombinant invasin nucleic acid operably linked to a polynucleotide that encodes a display polypeptide; (8) contacting the library of recombinant bacteriophage with a population of target cells; (9) removing unbound phage and phage which is bound to the surface of the target cells; and (10) recovering phage which are present within the target cells; and (11) repeating (6) through (10), as necessary, to obtain a further optimized recombinant delivery vehicle which exhibits further have enhanced ability to enter the target cells. In some embodiments the methods of evolving a bacteriophage-derived vaccine delivery vehicle to obtain a delivery vehicle having enhanced ability to enter a target cell can include the additional steps of (12) inserting into the optimized recombinant delivery vehicle a polynucleotide which encodes an antigen of interest, wherein the antigen of interest is expressed as a fusion polypeptide which comprises

a second display polypeptide; (13) administering the delivery vehicle to a test animal; and (14) determining whether the delivery vehicle is capable of inducing a CTL response in the test animal. Alternatively, the following steps can be employed: (12) inserting into the optimized recombinant delivery vehicle a polynucleotide which encodes an antigen of interest, wherein the antigen of interest is expressed as a fusion polypeptide which comprises a second display polypeptide; (13) administering the delivery vehicle to a test animal; and (14) determining whether the delivery vehicle is capable of inducing neutralizing antibodies against a pathogen which comprises the antigen of interest. An example of a target cell of interest for these methods is an antigen-presenting cell.

The present invention provides recombinant multivalent antigenic polypeptides that include a first antigenic determinant from a first disease-associated polypeptide and at least a second antigenic determinant from a second disease-associated polypeptide. The disease-associated polypeptides can be selected from the group consisting of cancer antigens, antigens associated with autoimmunity disorders, antigens associated with inflammatory conditions, antigens associated with allergic reactions, antigens associated with infectious agents, and other antigens that are associated with a disease condition.

In another embodiment, the invention provides a recombinant antigen library that contains recombinant nucleic acids that encode antigenic polypeptides. The libraries are typically obtained by reassembling (&/or subjecting to one or more directed evolution methods described herein), at least first and second forms of a nucleic acid which includes a polynucleotide sequence that encodes a disease-associated antigenic polypeptide, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant nucleic acids. Another embodiment of the invention provides methods of obtaining a polynucleotide that encodes a recombinant antigen having improved ability to induce an immune response to a disease condition. These methods involve: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which comprises a polynucleotide sequence that encodes an antigenic polypeptide that is associated with the disease condition, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant nucleic acids; and (2) screening the library to identify at least one optimized recombinant nucleic acid that encodes an optimized recombinant antigenic polypeptide that has improved ability to induce an immune response to the disease condition. These methods

optionally further involve: (3) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least one optimized recombinant nucleic acid with a further form of the nucleic acid, which is the same or different from the first and second forms, to produce a further library of recombinant nucleic acids; (4) screening the further library to identify at least one further optimized recombinant nucleic acid that encodes a polypeptide that has improved ability to induce an immune response to the disease condition; and (5) repeating (3) and (4), as necessary, until the further optimized recombinant nucleic acid encodes a polypeptide that has improved ability to induce an immune response to the disease condition. In some embodiments, the optimized recombinant nucleic acid encodes a multivalent antigenic polypeptide and the screening is accomplished by expressing the library of recombinant nucleic acids in a phage display expression vector such that the recombinant antigen is expressed as a fusion protein with a phage polypeptide that is displayed on a phage particle surface; contacting the phage with a first antibody that is specific for a first serotype of the pathogenic agent and selecting those phage that bind to the first antibody; and contacting those phage that bind to the first antibody with a second antibody that is specific for a second serotype of the pathogenic agent and selecting those phage that bind to the second antibody; wherein those phage that bind to the first antibody and the second antibody express a multivalent antigenic polypeptide.

The invention also provides methods of obtaining a recombinant viral vector which has an enhanced ability to induce an antiviral response in a cell. Methods of obtaining a recombinant genetic vaccine component that confers upon a genetic vaccine an enhanced ability to induce a desired immune response in a mammal

In additional embodiments, the invention provides methods of obtaining a recombinant genetic vaccine component that confers upon a genetic vaccine an enhanced ability to induce a desired immune response in a mammal. These methods involve: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which comprise a genetic vaccine vector, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant genetic vaccine vectors; (2) transfecting the library of recombinant vaccine vectors into a population of mammalian cells selected from the group consisting of peripheral blood T cells, T cell clones, freshly isolated monocytes/macrophages and dendritic cells; (3) staining the cells for the presence of one or more cytokines and

identifying cells which exhibit a cytokine staining pattern indicative of the desired immune response; and (4) obtaining recombinant vaccine vector nucleic acid sequences from the cells which exhibit the desired cytokine staining pattern.

Methods of improving the ability of a genetic vaccine vector to modulate an immune response

Also provided by the invention are methods of improving the ability of a genetic vaccine vector to modulate an immune response by: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which comprise a genetic vaccine vector, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant genetic vaccine vectors; (2) transfecting the library of recombinant genetic vaccine vectors into a population of antigen presenting cells; and (3) isolating from the cells optimized recombinant genetic vaccine vectors which exhibit enhanced ability to modulate a desired immune response.

Methods of obtaining a recombinant genetic vaccine vector that has an enhanced ability to induce a desired immune response in a mammal upon administration to the skin of the mammal.

Another embodiment of the invention provides methods of obtaining a recombinant genetic vaccine vector that has an enhanced ability to induce a desired immune response in a mammal upon administration to the skin of the mammal. These methods involve: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which comprise a genetic vaccine vector, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant genetic vaccine vectors; (2) topically applying the library of recombinant genetic vaccine vectors to skin of a mammal; (3) identifying vectors that induce an immune response; and (4) recovering genetic vaccine vectors from the skin cells which contain vectors that induce an immune response.

Methods of inducing an immune response in a mammal by topically applying to skin of the mammal a genetic vaccine vector, wherein the genetic vaccine vector is optimized for topical application through use of stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly.

The invention also provides methods of inducing an immune response in a mammal by topically applying to skin of the mammal a genetic vaccine vector, wherein the genetic vaccine vector is optimized for topical application through use of stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly. In some embodiments, the genetic vaccine is administered as a formulation selected from the group consisting of a transdermal patch, a cream, naked DNA, a mixture of DNA and a transfection-enhancing agent. Suitable transfection-enhancing agents include one or more agents selected from the group consisting of a lipid, a liposome, a protease, and a lipase. Alternatively, or in addition, the genetic vaccine can be administered after pretreatment of the skin by abrasion or hair removal.

Methods of obtaining an optimized genetic vaccine component that confers upon a genetic vaccine containing the component an enhanced ability to induce or inhibit apoptosis of a cell into which the vaccine is introduced.

In another embodiment, the invention provides methods of obtaining an optimized genetic vaccine component that confers upon a genetic vaccine containing the component an enhanced ability to induce or inhibit apoptosis of a cell into which the vaccine is introduced. These methods involve: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which comprise a nucleic acid that encodes an apoptosis- modulating polypeptide, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant nucleic acids; (2) transfecting the library of recombinant nucleic acids into a population of mammalian cells; (3) staining the cells for the presence of a cell membrane change which is indicative of apoptosis initiation; and (4) obtaining recombinant apoptosis-modulating genetic vaccine components from the cells which exhibit the desired apoptotic membrane changes.

Methods of obtaining a genetic vaccine component that confers upon a genetic vaccine reduced susceptibility to a CTL immune response in a host mammal.

Other embodiments of the invention provide methods of obtaining a genetic vaccine component that confers upon a genetic vaccine reduced susceptibility to a CTL immune response in a host mammal. These methods can involve: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which comprises a gene that encodes an inhibitor of a CTL immune response,

wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant CTL inhibitor nucleic acids; (2) introducing genetic vaccine vectors which comprise the library of recombinant CTL inhibitor nucleic acids into a plurality of human cells; (3) selecting cells which exhibit reduced MHC class I molecule expression; and (4) obtaining optimized recombinant CTL inhibitor nucleic acids from the selected cells.

Methods of obtaining a genetic vaccine component that confers upon a genetic vaccine reduced susceptibility to a CTL immune response in a host mammal.

The invention also provides methods of obtaining a genetic vaccine component that confers upon a genetic vaccine reduced susceptibility to a CTL immune response in a host mammal. These methods involve: (1) reassembling (&/or subjecting to one or more directed evolution methods described herein) at least first and second forms of a nucleic acid which comprises a gene that encodes an inhibitor of a CTL immune response, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant CTL inhibitor nucleic acids; (2) introducing viral vectors which comprise the library of recombinant CTL inhibitor nucleic acids into mammalian cells; (3) identifying mammalian cells which express a marker gene included in the viral vectors a predetermined time after introduction, wherein the identified cells are resistant to a CTL response; and (4) recovering as the genetic vaccine component the recombinant CTL inhibitor nucleic acids from the identified cells.

It is a general object of the invention to provide proteins and polypeptides that are derived from PfEMP1 proteins, nucleic acids encoding these proteins and antibodies that are specifically immunoreactive with these proteins. It is a further object to provide methods of using these various compositions in diagnosis, treatment or prevention of the onset of symptoms of a malaria parasite infection. It is a further object to provide methods of screening compounds to identify further compositions which may be used in these methods.

In one embodiment, the present invention provides substantially pure polypeptides which have amino acid sequences substantially homologous to the amino acid sequence of a PfEMP1 protein, or biologically active fragments thereof.

In alternative aspects, the polypeptides of the present invention are substantially homologous to the amino acid sequence shown, described &/or referenced herein (including

incorporated by reference), biologically active fragments or analogues thereof. Also provided are pharmaceutical compositions comprising these polypeptides.

In another embodiment, the present invention provides nucleic acids which encode the above-described polypeptides. Exemplary nucleic acids of the invention can be substantially homologous to a part or whole of the nucleic acid sequence shown, described &/or referenced herein (including incorporated by reference) or the nucleic acid encoding for the sequences shown, described &/or referenced herein (including incorporated by reference). The present invention also provides expression vectors comprising these nucleic acid sequences and cells capable of expressing same.

In an additional embodiment, the present invention provides antibodies which recognize and bind PfEMP1 polypeptides or biologically active fragments thereof. These peptides can recognize and bind PfEMP1 proteins associated with infection by more than one variant of P. falciparum. In a further embodiment, the present invention provides methods of inhibiting the formation of PfEMP1/ligand complex, comprising contacting PfEMP1 or its ligands with polypeptides of the present invention. In a related embodiment, the present invention provides methods of inhibiting sequestration of erythrocytes in a patient suffering from a malaria infection, comprising administering to said patient, an effective amount of a polypeptide of the present invention. such administration may be carried out prior to or following infection. In still another embodiment, the present invention provides a method of detecting the presence or absence of PfEMP1 in a sample. The method comprises exposing the sample to an antibody of the invention, and detecting binding, if any, between the antibody and a component of the sample. In an additional embodiment, the present invention provides a method of determining whether a test compound is an antagonist of PfEMP1/ligand complex formation. The method comprises incubating the test compound with PfEMP1 or a biologically active fragment thereof, and its ligand, under conditions which permit the formation of the complex. The amount of complex formed in the presence of the test compound is determined and compared with the amount of complex formed in the absence of the test compound. A decrease in the amount of complex formed in the presence of the test compound is indicative that the compound is an antagonist of PfEMP1/ligand complex formation.

<u>Summary of Directed Evolution Approaches</u>

This invention also relates generally to the field of nucleic acid engineering and correspondingly encoded recombinant protein engineering. More particularly, the invention relates to the directed evolution of nucleic acids and screening of clones containing the evolved nucleic acids for resultant activity(ies) of interest, such nucleic acid activity(ies) &/or specified protein, particularly enzyme, activity(ies) of interest. Mutagenized molecules provided by this invention may have chimeric molecules and molecules with point mutations, including biological molecules that contain a carbohydrate, a lipid, a nucleic acid, &/or a protein component, and specific but non-limiting examples of these include antibiotics, antibodies, enzymes, and steroidal and non-steroidal hormones. This invention relates generally to a method of: 1) preparing a progeny generation of molecule(s) (including a molecule that is comprised of a polynucleotide sequence, a molecule that is comprised of a polypeptide sequence, and a molecules that is comprised in part of a polynucleotide sequence and in part of a polypeptide sequence), that is mutagenized to achieve at least one point mutation, addition, deletion, &/or chimerization, from one or more ancestral or parental generation template(s); 2) screening the progeny generation molecule(s) – in one aspect, using a high throughput method - for at least one property of interest (such as an improvement in an enzyme activity or an increase in stability or a novel chemotherapeutic effect); 3) optionally obtaining &/or cataloguing structural &/or and functional information regarding the parental &/or progeny generation molecules; and 4) optionally repeating any of steps 1) to 3).

In a one embodiment, there is generated (e.g. from a parent polynucleotide template) - in what is termed "codon site-saturation mutagenesis" - a progeny generation of polynucleotides, each having at least one set of up to three contiguous point mutations (i.e. different bases comprising a new codon), such that every codon (or every family of degenerate codons encoding the same amino acid) is represented at each codon position. Corresponding to - and encoded by - this progeny generation of polynucleotides, there is also generated a set of progeny polypeptides, each having at least one single amino acid point mutation. In a one aspect, there is generated - in what is termed "amino acid site-saturation mutagenesis" - one such mutant polypeptide for each of the 19 naturally encoded polypeptide-forming alpha-amino acid substitutions at each and every amino acid position along the polypeptide. This yields - for each and every amino acid position along the parental polypeptide - a total of 20 distinct progeny polypeptides including the original

amino acid, or potentially more than 21 distinct progeny polypeptides if additional amino acids are used either instead of or in addition to the 20 naturally encoded amino acids. Thus, in another aspect, this approach is also serviceable for generating mutants containing - in addition to &/or in combination with the 20 naturally encoded polypeptide-forming alpha-amino acids - other rare &/or not naturally-encoded amino acids and amino acid derivatives. In yet another aspect, this approach is also serviceable for generating mutants by the use of - in addition to &/or in combination with natural or unaltered codon recognition systems of suitable hosts - altered, mutagenized, &/or designer codon recognition systems (such as in a host cell with one or more altered tRNA molecules).

In yet another aspect, this invention relates to recombination and more specifically to a method for preparing polynucleotides encoding a polypeptide by a method of *in vivo* re-assortment of polynucleotide sequences containing regions of partial homology, assembling the polynucleotides to form at least one polynucleotide and screening the polynucleotides for the production of polypeptide(s) having a useful property.

In one embodiment, this invention is serviceable for analyzing and cataloguing - with respect to any molecular property (e.g. an enzymatic activity) or combination of properties allowed by current technology - the effects of any mutational change achieved (including particularly saturation mutagenesis). Thus, a comprehensive method is provided for determining the effect of changing each amino acid in a parental polypeptide into each of at least 19 possible substitutions. This allows each amino acid in a parental polypeptide to be characterized and catalogued according to its spectrum of potential effects on a measurable property of the polypeptide. In another aspect, the method of the present invention utilizes the natural property of cells to recombine molecules and/or to mediate reductive processes that reduce the complexity of sequences and extent of repeated or consecutive sequences possessing regions of homology.

It is an object of the present invention to provide a method for generating hybrid polynucleotides encoding biologically active hybrid polypeptides with enhanced activities. In accomplishing these and other objects, there has been provided, in accordance with one aspect of the invention, a method for introducing polynucleotides into a suitable host cell and growing the host cell under conditions that produce a hybrid polynucleotide.

In another aspect of the invention, the invention provides a method for screening for biologically active hybrid polypeptides encoded by hybrid polynucleotides. The present

method allows for the identification of biologically active hybrid polypeptides with enhanced biological activities.

Methods for Determining the Immunogenicity of a Test Molecule Using Immunocompromised Mammals Reconstituted with Human Lymphocytes

The invention provides a method for determining the immunogenicity of a test molecule (i.e., a test antigen) comprising the following steps: (a) providing an immunocompromised non-human mammal populated with a plurality of human lymphocytes; (b) providing a test molecule; (c) administering the test molecule to the immunocompromised non-human mammal of step (a); (d) determining the test molecule-specific immune response of the human lymphocytes; and, (e) removing a sample of human lymphocytes from the non-human mammal and testing for their ability to proliferate or produce antibodies in response to challenge by the test molecule. No response or a diminished response (e.g., generating antibodies with a Kd of less than about 10^{-6}) would be indicative of low immunogenicity of the test molecule. The immunocompromised non-human mammal can be any mammal, e.g., a SCID mouse or rat. The non-human mammals can be genetically manipulated to be immuno-compromised (e.g., SCID) or they can be treated with chemicals (drugs) and/or irradiated.

The antigen structure or dosage, the route and/or number of administrations, the formulation (e.g., the adjuvant) and/or the non-human animal can be varied and/or manipulated to generate the desired immune response ("immunogenicity"), e.g., the form of response (e.g., humoral or cellular response), isotype of response (e.g., a humoral IgM, IgG, IgA, IgE, IgD, or a cellular T helper, T killer or T suppressor cell response), affinity of resultant antibody (e.g., high affinity (e.g., about 10^6 or higher) or low affinity), and the like. For example, the test molecule can be administered two or more times before determining the results of the test molecule-specific immune response, e.g., nature of response, affinity of antibodies, and the like. The test molecule can administered two or more times and the resultant immune response can be determined after each administration. Alternatively, the test molecule can be modified between each round of administration and re-testing of immune response.

In alternative aspects, the test molecule comprises a polypeptide, a peptide, a lipid, a nucleic acid, a small molecule and/or a polysaccharide. The polypeptide can be synthetic, isolated from a natural source or recombinant.

In alternative aspects, the test molecule is structurally modified after each or one or more administrations. The process can be reiterated to generate a desired response. For example, after an initial administration, if the response generate a low humoral response and a high humoral response is desired, the test molecule is structurally modified, re-administered and the resultant immune response is analyzed. In another example, if a T helper response is generated and a T suppressor response is desired, the test molecule is structurally modified, re-administered and retested. This process can be reiterated as many times as necessary to generate a desired response. The structural modification in the test molecule can be combined with other changes in administration or formulation, e.g., dosages, routes of administration and the like.

If the test molecule is a polypeptide (including peptides), it can be modified from its native (e.g., wild type) sequence by modifications, additions or deletions. The modifications can be a change in amino acid residue(s) (e.g., either a conservative change, such as a hydrophobic residue to another hydrophobic residue, or a non-conservative change, e.g., a hydrophobic residue to a hydrophilic residue) or a change in the structure of a residue, e.g., a post-translational change (e.g., phosphorylation, lipidation) or a post-synthetic structural modification in an amino acid residue (e.g., to a cyclodepsipeptide, mycosporine-like amino acid, amidation, oxidation and the like). Modifications in the test polypeptide can be introduced by, e.g., error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, gene site saturated mutagenesis (GSSM), synthetic ligation reassembly (SLR) and/or a combination thereof. In alternative aspects, the modifications, additions or deletions are introduced by, e.g., recombination, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation and/or a combination thereof.

The plurality of human lymphocytes can comprise human peripheral blood lymphocytes. The lymphocytes can be unchallenged ("naïve"), pre-challenged (antigen

stimulated) or activated (e.g., mitogen-, hormone- or interleukin-activated) cells. The immune response can comprise a humoral response (an antibody based response) or a cellular (white blood cell) response. In one aspect, the isotypes of the antibodies generated in the humoral response are characterized.

The human lymphocyte can be, e.g., a sample of human lymphocytes comprising T cells, macrophages, monocytes, dendritic cells, B cells and/or plasma cells.

Methods for Generating High Affinity Antibodies

The invention provides methods for generating high affinity antibodies comprising the following steps: (a) providing a sample of isolated B lymphocytes; (b) isolating or cloning from the isolated B lymphocytes a nucleic acid encoding an antibody molecule; (c) translating the antibody molecule-encoding nucleic acid and placing the translated polypeptides in conditions wherein VH/VL pairing can occur to form an antigen binding molecule; (d) screening the antigen binding molecule for its ability to selectively bind to an antigen and its affinity for the antigen; (e) isolating the antigen binding molecule-encoding nucleic acid and changing its nucleic acid sequence; and, (f) re-screening the antigen binding molecule for its ability to selectively bind to an antigen by (i) expressing the antibody-encoding nucleic acid isolated in step (e) to generate antigen binding polypeptides, (ii) placing the expressed polypeptides in conditions wherein VH/VL pairing can occur to form antigen binding molecules, and, (iii) screening the antigen binding molecules for their ability to selectively bind to the antigen and having an antigen binding affinity higher than the antigen binding molecule screened in step (d).

In alternative aspects, the B lymphocytes are human or mouse B lymphocytes. The B lymphocytes can be isolated by FACS sorting. The B lymphocytes can be labeled with fluorescent tags before the FACS sorting.

In one aspect, the nucleic acid encoding the antibody comprises an mRNA. The nucleic acid encoding an antibody can be isolated by RT-PCR.

In one aspect, the B lymphocytes are pooled into separate fractions before the antibody-encoding nucleic acid is isolated or cloned. The B lymphocytes can be pooled into separate fractions of about 1000 cells, 500 cells, 100 cells, 50 cells, 25 cells or 10 cells per fraction.

In alternative aspects, the nucleic acid sequences is changed by mutagenesis, base residue insertion or base residue deletion. Evolution technologies can be used to further engineer these sequences, including, e.g., Gene Site Saturation MutagenesisTM (GSSM) and

GeneReassemblyTM (Diversa Corporation, San Diego, CA), as described in further detail herein. Alternatively, the nucleic acid sequences can be changed or "evolved" or "genetically engineered" by, e.g., error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, in vivo mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis, site-specific mutagenesis, gene reassembly, gene site saturated mutagenesis (GSSM), synthetic ligation reassembly (SLR) and/or a combination thereof. In alternative aspects, the modifications, additions or deletions are introduced by, e.g., recombination, recursive sequence recombination, phosphothioate-modified DNA mutagenesis, uracil-containing template mutagenesis, gapped duplex mutagenesis, point mismatch repair mutagenesis, repair-deficient host strain mutagenesis, chemical mutagenesis, radiogenic mutagenesis, deletion mutagenesis, restriction-selection mutagenesis, restriction-purification mutagenesis, artificial gene synthesis, ensemble mutagenesis, chimeric nucleic acid multimer creation and/or a combination thereof. In one aspect, these methods are iteratively repeated until an antibody having an altered or different activity or an altered or different stability from that of the antibody to be "evolved" is produced. In one aspect, the CDR3 region of the antigen binding molecule-encoding nucleic acid sequence is changed or "evolved."

Antibody Arrays and Methods of Making and Using Them

The invention provides an array (e.g., biochip) comprising a plurality of polypeptides, wherein each polypeptide is immobilized to a discrete and known spot on a substrate surface to form an array of polypeptides, wherein the plurality of polypeptides comprise a sample of (i.e., a subset of), or all of, the antigen binding sites that are isolated from and/or expressed by an individual, or, complementary to antigen binding sites isolated from and/or expressed by the individual. In one aspect, one or more of these antigen binding sites can be an antigen binding site encoded by a nucleic acid modified, or "evolved," by one or more of the methods of the invention, as described herein. In one aspect, one or more of these antigen binding sites can be an antigen binding site encoded by a nucleic acid from a library of the invention (e.g., antigen binding sites encoded by a library of nucleic acids).

In one aspect, the polypeptides on the array comprise antigen binding sites isolated from or complementary to antigen binding sites of antibodies expressed by the individual, including secreted or cell-expressed (e.g., cell-bound) antibodies or fragments thereof. In

one aspect, antigen binding sites can be isolated from or complementary to antigen binding sites expressed on circulating antibodies expressed by the individual.

One or more of these secreted, circulating and/or cell-expressed antigen binding sites can be encoded by a nucleic acid modified, or "evolved," by one or more of the methods of the invention, as described herein.

In one aspect, the cell-bound antibodies comprise B cell-bound, plasma cell-bound or macrophage-bound antibodies. The cell-bound antibodies can be IgG, IgM, IgD, IgA and/or IgE. In one aspect, the sample comprises antigen binding sites isolated from or complementary to antigen binding sites expressed on cell-bound and circulating antibodies expressed by the individual. In one aspect, the sample comprises a complete repertoire of the antigen binding sites of antibodies expressed by the individual.

In one aspect, the antigen binding site comprises a polypeptide selected from the group consisting of a single stranded antigen binding polypeptide, a Fab fragment, an Fc fragment, a F(ab')2 fragment, a Fv fragment and a complementarity determining region (CDR). The antigen binding site can comprise an antibody polypeptide comprising two light chains and two heavy chains.

In one aspect, the sample comprises a complement of antigen binding sites isolated from or complementary to antigen binding sites expressed in a lymph node of the individual. The lymph node can be isolated by, e.g., dissection or biopsy. The cells can be harvested by aspiration or by cell sorting.

In one aspect, the sample comprises a complement of antigen binding sites isolated from or complementary to T cell receptors (TCRs) expressed by the individual. The sample can comprise a complement of antigen binding sites isolated from or complementary to T cell receptors (TCRs) and antibodies expressed by the individual. The sample can comprise a complete repertoire of the T cell receptors (TCRs) expressed in the individual.

The individual can be any mammal, e.g., a mouse, a rat or a human.

The plurality of polypeptides can further comprise a sample comprising antigen binding sites that are structural variations of antigen binding sites expressed by the individual. The structural variations can be made by a method comprising the following steps: (a) providing a template polynucleotide, wherein the template polynucleotide comprises sequence encoding an antigen binding site; (b) providing a plurality of oligonucleotides, wherein each oligonucleotide comprises a sequence homologous to the

template polynucleotide, thereby targeting a specific sequence of the template polynucleotide, and a sequence that is a variant of antigen binding site-encoding sequence; (c) generating progeny polynucleotides comprising non-stochastic sequence variations by replicating the template polynucleotide of step (a) with the oligonucleotides of step (b), thereby generating polynucleotides comprising antigen binding site-encoding sequence variations; and, (d) expressing the polynucleotides to generate polypeptides comprising antigen binding sites that are structural variations of antigen binding sites expressed by the individual. The sequence homologous to the template polynucleotide can be x bases long, wherein x is an integer between 10 and 30, or, between 2 and 20. The oligonucleotide of step (b) can further comprises a second sequence homologous to the template polynucleotide and the variant sequence is flanked by the sequences homologous to the template polynucleotide. A codon encoding an amino acid in the antigen binding site can be targeted to be modified, and the plurality of oligonucleotides comprise variant sequences encoding all nineteen naturally-occurring amino acid variants for the targeted codon, thereby generating an antigen binding site polypeptide for all nineteen possible natural amino acid variations at the targeted amino acid. In one apect, codons encoding all amino acids in the antigen binding site are targeted to be modified. The plurality of oligonucleotides can comprise variant sequences encoding all nineteen naturally-occurring amino acid variants for the targeted codon, thereby generating an antigen binding site polypeptide for all nineteen possible natural amino acid variations at each targeted amino acid. An oligonucleotide of step (b) can further comprise a nucleic acid sequence capable of introducing one or more nucleotide residues into the template polynucleotide, or, deleting one or more residue from the template polynucleotide.

Structural variations of antigen binding sites can be made by a method comprising the following steps: (a) providing a template polynucleotide, wherein the template polynucleotide comprises sequence encoding an antigen binding site; (b) providing a plurality of building block polynucleotides, wherein the building block polynucleotides are designed to cross-over reassemble with the template polynucleotide at a predetermined sequence, and a building block polynucleotide comprises a sequence that is a variant of an antigen binding site-encoding sequence and a sequence homologous to the template polynucleotide flanking the variant sequence; (c) combining a block polynucleotide with a template polynucleotide such that the building block polynucleotide cross-over reassembles with the template polynucleotide to generate polynucleotides comprising antigen binding

site-encoding sequence variations; and (d) expressing the polynucleotides to generate polypeptides comprising antigen binding sites that are structural variations of antigen binding sites expressed by the individual.

In one aspect, the building block polynucleotides comprise a sequence homologous to the template polynucleotide x bases long, wherein x is an integer between 10 and 30. The building block polynucleotides can comprise a sequence that is a variant of the template polynucleotide x bases long, wherein x is an integer between 2 and 20. The codon encoding an amino acid in the antigen binding site can be targeted to be modified, and the building block polynucleotides comprise variant sequences encoding all nineteen naturally-occurring amino acid variants for the targeted codon, thereby generating an antigen binding site polypeptide for all nineteen possible natural amino acid variations at the targeted amino acid. The codons encoding all amino acids in the antigen binding site can be targeted to be modified.

In one aspect, the plurality of oligonucleotides comprise variant sequences encoding all nineteen naturally-occurring amino acid variants for the targeted codon, thereby generating an antigen binding site polypeptide for all nineteen possible natural amino acid variations at each targeted amino acid. The building block polynucleotide can further comprise a nucleic acid sequence capable of introducing one or more nucleotide residues into the template polynucleotide, or, deleting one or more residue from the template polynucleotide. In one aspect, a variant antigen binding site has a higher affinity for antigen than the template antigen binding site.

In one aspect, the methods for modifying antigen binding site structures can further comprise iteratively repeating steps (a) through (d), thereby generating further structural variations of antigen binding sites. In one aspect, the methods further comprising selecting a variant antigen binding site capable of enzymatically catalyzing a reaction.

In invention provides methods of making arrays comprising a plurality of polypeptide antigen binding sites, the methods comprising the following steps: (a) providing a plurality of polypeptides comprising a sample (e.g., a subset) of antigen binding sites that are isolated from or complementary to antigen binding sites expressed by an individual; and, (b) immobilizing to a discrete and known spot on a substrate surface one or more polypeptides each comprising the same antigen binding site, thereby forming an array of antigen binding site polypeptides. In one aspect, the sample comprises antigen binding sites isolated from or

complementary to antigen binding sites expressed on secreted antibodies expressed by the individual. The sample can comprise antigen binding sites isolated from or complementary to antigen binding sites expressed on circulating antibodies expressed by the individual. The sample can comprise antigen binding sites isolated from or complementary to antigen binding sites expressed on cell-bound antibodies expressed by the individual. In one aspect, the cell-bound antibodies comprise B cell-bound antibodies. The sample can comprise a complement of antigen binding sites isolated from or complementary to antigen binding sites expressed in a lymph node of the individual. The sample can comprise antigen binding sites isolated from or complementary to antigen binding sites expressed on cell-bound and circulating antibodies expressed by the individual. The sample can comprise a complete repertoire of the antigen binding sites of antibodies expressed by the individual.

In one aspect, the sample comprises a complement of antigen binding sites isolated from or complementary to T cell receptors (TCRs) expressed by the individual. The sample can comprise a complement of antigen binding sites isolated from or complementary to T cell receptors (TCRs) and antibodies expressed by the individual.

In one aspect, the sample comprises a complete repertoire of the antigen binding sites expressed in the individual. The antigen binding sites can comprise antibodies comprising a μ, γ, $\gamma2$, $\gamma3$, $\gamma4$, δ, ε, $\alpha1$ or $\alpha2$ constant region.

In one aspect, the antigen binding sites are generated by expression of nucleic acid generated by amplification of nucleic acid from the individual. The amplification can comprise, e.g., polymerase chain reaction (PCR). The nucleic acid can comprise a cDNA library. The cDNA library can be made from nucleic acid isolated from B cells or plasma cells. The cDNA library can be made from nucleic acid isolated from, e.g., a lymph node, a spleen, a thymus, B cells or plasma cells. The antigen binding sites and/or the nucleic acid encoding them can be isolated from, e.g., a lymph node, a spleen, a thymus, a blood or serum sample or a biopsy.

The invention provides methods of selecting an antibody capable of selectively binding to an antigen, the methods comprising the following steps: (a) providing an array comprising a plurality of polypeptides, wherein each polypeptide is immobilized to a discrete and known spot on a substrate surface to form an array of polypeptides, wherein the plurality of polypeptides comprise a sample (e.g., a subset) of antigen binding sites expressed by an individual; (b) contacting the array with an antigen under conditions where the antigen can

specifically bind to the antibody; (c) washing unbound antigen off the array; and, (c) determining which spot has selectively bound the antigen, thereby selecting an antibody capable of selectively binding to the antigen.. In one apect, the antigen is contacted with the array under varying conditions of increasingly stringent conditions, selecting an antibody having a high affinity to the antigen. The affinity can be selected from the group consisting of about 1 x 105 M-1, about $1 \times 10^5 M^{-1}$, about $1 \times 10^6 M^{-1}$, about $1 \times 10^7 M^{-1}$, about $1 \times 10^8 M^{-1}$, about $1 \times 10^9 M^{-1}$, about $2 \times 10^9 M^{-1}$, about $5 \times 10^9 M^{-1}$, about $1 \times 10^{10} M^{-1}$, about $1 \times 10^{11} M^{-1}$ and greater than $1 \times 10^{11} M^{-1}$. In one apect, the antigen comprises a detectable label, such as a fluorescent molecule, e.g., umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin.

In alternative apects, the detectable label comprises a radioactive molecule or an enzyme, such as a horseradish peroxidase, beta-galactosidase, luciferase or an alkaline phosphatase.

The details of one or more aspects of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

All publications, GenBank Accession references (sequences), ATCC Deposits, patents and patent applications cited herein are hereby expressly incorporated by reference for all purposes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Exonuclease Activity. Figure 1 shows the activity of the enzyme exonuclease III. This is an exemplary enzyme that can be used to shuffle, assemble, reassemble, recombine, and/or concatenate polynucleotide building blocks. The asterisk indicates that the enzyme acts from the 3' direction towards the 5' direction of the polynucleotide substrate.

Figure 2. Generation of A Nucleic Acid Building Block by Polymerase-Based Amplification. Figure 2 illustrates a method of generating a double-stranded nucleic acid building block with two overhangs using a polymerase-based amplification reaction (e.g., PCR). As illustrated, a first polymerase-based amplification reaction using a first set of primers, F_2 and R_1, is used to generate a blunt-ended product (labeled Reaction 1, Product 1), which is essentially identical to Product A. A second polymerase-based amplification

reaction using a second set of primers, F_1 and R_2, is used to generate a blunt-ended product (labeled Reaction 2, Product 2), which is essentially identical to Product B. These two products are then mixed and allowed to melt and anneal, generating a potentially useful double-stranded nucleic acid building block with two overhangs. In the example of Fig. 1, the product with the 3' overhangs (Product C) is selected for by nuclease-based degradation of the other 3 products using a 3' acting exonuclease, such as exonuclease III. Alternate primers are shown in parenthesis to illustrate serviceable primers may overlap, and additionally that serviceable primers may be of different lengths, as shown.

Figure 3. Unique Overhangs And Unique Couplings. Figure 3 illustrates the point that the number of unique overhangs of each size (e.g. the total number of unique overhangs composed of 1 or 2 or 3, etc. nucleotides) exceeds the number of unique couplings that can result from the use of all the unique overhangs of that size. For example, there are 4 unique 3' overhangs composed of a single nucleotide, and 4 unique 5' overhangs composed of a single nucleotide. Yet the total number of unique couplings that can be made using all the 8 unique single-nucleotide 3' overhangs and single-nucleotide 5' overhangs is 4.

Figure 4. Unique Overall Assembly Order Achieved by Sequentially Coupling the Building Blocks

Figure 4 illustrates the fact that in order to assemble a total of "n" nucleic acid building blocks, "n-1" couplings are needed. Yet it is sometimes the case that the number of unique couplings available for use is fewer that the "n-1" value. Under these, and other, circumstances a stringent non-stochastic overall assembly order can still be achieved by performing the assembly process in sequential steps. In this example, 2 sequential steps are used to achieve a designed overall assembly order for five nucleic acid building blocks. In this illustration the designed overall assembly order for the five nucleic acid building blocks is: 5'-(#1-#2-#3-#4-#5)-3', where #1 represents building block number 1, etc.

Figure 5. Unique Couplings Available Using a Two-Nucleotide 3' Overhang. Figure 5 further illustrates the point that the number of unique overhangs of each size (here, e.g. the total number of unique overhangs composed of 2 nucleotides) exceeds the number of unique couplings that can result from the use of all the unique overhangs of that size. For example, there are 16 unique 3' overhangs composed of two nucleotides, and another 16 unique 5' overhangs composed of two nucleotides, for a total of 32 as shown. Yet the total number of couplings that are unique and not self-binding that can be made using all the 32 unique

double-nucleotide 3' overhangs and double-nucleotide 5' overhangs is 12. Some apparently unique couplings have "identical twins" (marked in the same shading), which are visually obvious in this illustration. Still other overhangs contain nucleotide sequences that can self-bind in a palindromic fashion, as shown and labeled in this figure; thus they not contribute the high stringency to the overall assembly order.

Figure 6. Generation of an Exhaustive Set of Chimeric Combinations by Synthetic Ligation Reassembly. Figure 6 showcases the power of this invention in its ability to generate exhaustively and systematically all possible combinations of the nucleic acid building blocks designed in this example. Particularly large sets (or libraries) of progeny chimeric molecules can be generated. Because this method can be performed exhaustively and systematically, the method application can be repeated by choosing new demarcation points and with correspondingly newly designed nucleic acid building blocks, bypassing the burden of re-generating and re-screening previously examined and rejected molecular species. It is appreciated that, codon wobble can be used to advantage to increase the frequency of a demarcation point. In other words, a particular base can often be substituted into a nucleic acid building block without altering the amino acid encoded by progenitor codon (that is now altered codon) because of codon degeneracy. As illustrated, demarcation points are chosen upon alignment of 8 progenitor templates. Nucleic acid building blocks including their overhangs (which are serviceable for the formation of ordered couplings) are then designed and synthesized. In this instance, 18 nucleic acid building blocks are generated based on the sequence of each of the 8 progenitor templates, for a total of 144 nucleic acid building blocks (or double-stranded oligos). Performing the ligation synthesis procedure will then produce a library of progeny molecules comprised of yield of 8^{18} (or over 1.8×10^{16}) chimeras.

Figure 7. Synthetic genes from oligos:. According to one embodiment of this invention, double-stranded nucleic acid building blocks are designed by aligning a plurality of progenitor nucleic acid templates. In one aspect, these templates contain some homology and some heterology. The nucleic acids may encode related proteins, such as related enzymes, which relationship may be based on function or structure or both. Figure 7 shows the alignment of three polynucleotide progenitor templates and the selection of demarcation points (boxed) shared by all the progenitor molecules. In this particular example, the nucleic

acid building blocks derived from each of the progenitor templates were chosen to be approximately 30 to 50 nucleotides in length.

Figure 8. Nucleic acid building blocks for synthetic ligation gene reassembly. Figure 8 shows the nucleic acid building blocks from the example in Figure 7. The nucleic acid building blocks are shown here in generic cartoon form, with their compatible overhangs, including both 5' and 3' overhangs. There are 22 total nucleic acid building blocks derived from each of the 3 progenitor templates. Thus, the ligation synthesis procedure can produce a library of progeny molecules comprised of yield of 3^{22} (or over 3.1×10^{10}) chimeras.

Figure 9. Addition of Introns by Synthetic Ligation Reassembly. Figure 9 shows in generic cartoon form that an intron may be introduced into a chimeric progeny molecule by way of a nucleic acid building block. It is appreciated that introns often have consensus sequences at both termini in order to render them operational. It is also appreciated that, in addition to enabling gene splicing, introns may serve an additional purpose by providing sites of homology to other nucleic acids to enable homologous recombination. For this purpose, and potentially others, it may be sometimes desirable to generate a large nucleic acid building block for introducing an intron. If the size is overly large easily generating by direct chemical synthesis of two single stranded oligos, such a specialized nucleic acid building block may also be generated by direct chemical synthesis of more than two single stranded oligos or by using a polymerase-based amplification reaction as shown, described &/or referenced herein (including incorporated by reference).

Figure 10. Ligation Reassembly Using Fewer Than All The Nucleotides Of An Overhang. Figure 10 shows that coupling can occur in a manner that does not make use of every nucleotide in a participating overhang. The coupling is particularly lively to survive (e.g. in a transformed host) if the coupling reinforced by treatment with a ligase enzyme to form what may be referred to as a "gap ligation" or a "gapped ligation". It is appreciated that, as shown, this type of coupling can contribute to generation of unwanted background product(s), but it can also be used advantageously increase the diversity of the progeny library generated by the designed ligation reassembly.

Figure 11. Avoidance of unwanted self-ligation in palindromic couplings. As mentioned before and shown, described &/or referenced herein (including incorporated by reference), certain overhangs are able to undergo self-coupling to form a palindromic coupling. A coupling is strengthened substantially if it is reinforced by treatment with a

ligase enzyme. Accordingly, it is appreciated that the lack of 5' phosphates on these overhangs, as shown, can be used advantageously to prevent this type of palindromic self-ligation. Accordingly, this invention provides that nucleic acid building blocks can be chemically made (or ordered) that lack a 5' phosphate group (or alternatively they can be remove – e.g. by treatment with a phosphatase enzyme such as a calf intestinal alkaline phosphatase (CIAP) – in order to prevent palindromic self-ligations in ligation reassembly processes.

Figure 12. Site-directed mutagenesis by polymerase-based extension. Panel A. This figure shows one method of site-directed mutagenesis, among many methods of site-directed mutagenesis, that are serviceable for performing site-saturation mutagenesis. Section (1) shows the first and second mutagenic primer annealed to a circular closed double-stranded plasmid. The dot and the open-sided triangle indicate the mutagenic sites in the mutagenic primers. The arrows indicate the direction of synthesis. Section (2) shows the newly synthesized (mutagenized) DNA strands annealed to each other. The parental DNA can be treated with a selection enzyme. The mutagenized DNA strands are shown as being annealed to form a double-stranded mutagenized circular DNA intermediate. The dot and the open-sided triangle indicate the mutagenic sites in the experimentally generated progeny (mutagenized) DNA strands. Note that the staggered openings on the mutagenized DNA strands form "sticky" ends. Section (3) shows the first and second mutagenic primer annealed to the mutagenized DNA strands of Section (2). The arrows indicate the direction of synthesis. Note the opening on each of the mutagenized DNA strands (i.e. they have not been ligated). Section (4) shows a "Gapped Product", which is composed of second generation mutagenized DNA strands, synthesized using the mutagenized DNA strands (shown in Step (2)) as a template. The DNA strands of the "Gapped Product" are shown as being annealed to form a double-stranded mutagenized circular DNA intermediate. The dot and the open-sided triangle indicate the mutagenic sites in the mutagenized DNA strands. Note the large gap in each of the mutagenized DNA strands. Section (5) shows the "Gapped Product" annealed to the parental (non-mutated) plasmid, enabling polymerase-based synthesis to occur. The arrows indicate the direction of synthesis. Section (6) shows the newly synthesized DNA strands, as being annealed to form a double-stranded mutagenized circular DNA product. The dot and the open-sided triangle indicate the mutagenic sites in the mutagenized DNA strands. Note the staggered openings on the mutagenized DNA

strands. Also note the presence of both mutagenic sites on each of the mutagenized DNA strands.

Panel B. This figure shows two possible molecular structures produced from the amplification steps of Figure 12A. Molecule (A) is shown also in Section (2) of Figure 12A. Molecule (B) is also shown in Section (6) of Figure 12A.

Figure 13. Site-directed mutagenesis by polymerase-based extension and ligase-based ligation. Panel A. This figure shows one method of site-directed mutagenesis, among many methods of site-directed mutagenesis, that are serviceable for performing site-saturation mutagenesis. Section (1) shows the first and second mutagenic primer annealed to a circular closed double-stranded plasmid. The dot and the open-sided triangle indicate the mutagenic sites in the mutagenic primers. The arrows indicate the direction of synthesis. Section (2) shows the newly synthesized (mutagenized) DNA strands annealed to each other. The parental DNA can be treated with a selection enzyme. The mutagenized DNA strands are shown as being annealed to form a double-stranded mutagenized circular DNA intermediate. The dot and the open-sided triangle indicate the mutagenic sites in the experimentally generated progeny (mutagenized) DNA strands. Note that the staggered openings on the mutagenized DNA strands form "sticky" ends. Section (3) shows the resultant double-stranded mutagenized circular DNA molecule produced after the double-stranded mutagenized circular DNA intermediate of Section (2) is ligated (e.g. with T4 DNA ligase). Section (4) shows the first and second mutagenic primer annealed to the mutagenized DNA strands of Section (3). The arrows indicate the direction of synthesis. Section (5) shows the recently generated (blue) mutagenized DNA strands as being annealed to form a double-stranded mutagenized circular DNA intermediate. The dot and the open-sided triangle indicate the mutagenic sites in the recently generated mutagenized DNA strands (blue). Note that the staggered openings on the mutagenized DNA strands form "sticky ends". Also note the presence of both mutagenic sites on each of the two recently generated mutagenized DNA strands (blue). Note the opening on each of the mutagenized DNA strands (i.e. they have not been ligated). Section (6) shows the resultant double-stranded mutagenized circular DNA molecule produced after the double-stranded mutagenized circular DNA intermediate of Section (5) is ligated (e.g. using T4 DNA ligase). The dot and the open-sided triangle indicate the mutagenic sites in the mutagenized DNA

molecules. Again, note the presence of both mutagenic sites on each of the mutagenized DNA strands.

Panel B. This figure shows two molecular structures produced from the amplification steps of Figure 13A. Molecule (A) is also shown in Section (3) of Figure 13A. Molecule (B) is produced in Section (6) of Figure 13A.

Figure 14: Strategy for Obtaining and Using Nucleic Acid Binding Proteins that Facilitate Entry of Genetic Vaccines. Shown here is a strategy for obtaining and using nucleic acid binding proteins that facilitate entry of genetic vaccines, in particular, naked DNA, into target cells. Members of a library obtained by the directed evolution methods described herein are linked to a coding region of M 13 protein VIII so that a fusion protein is displayed on the surface of the phage particles. Phage that efficiently enter the desired target tissue are identified, and the fusion protein is then used to coat a genetic vaccine nucleic acid.

Figure 15: A schematic representation of a method for generating a chimeric, multivalent antigen that has immunogenic regions from multiple antigens. Antibodies to each of the non-chimeric parental immunogenic polypeptides are specific for the respective organisms (A, B, C). After carrying out the directed evolution and selection methods of the invention, however, a chimeric immunogenic polypeptide is obtained that is recognized by antibodies raised against each of the three parental immunogenic polypeptides.

Figure 16A and Figure 16B: Method for Obtaining Non-Stochastically Generated Polypeptides that can induce a Broad-Spectrum Immune Response. Shown here is a schematic for a method by which one can obtain non-stochastically generated polypeptides that can induce a broad-spectrum immune response. In Figure 16A, wild-type immunogenic polypeptides from the pathogens A, B, and C provide protection against the corresponding pathogen from which the polypeptide is derived, but little or no cross-protection against the other pathogens (left panel). After evolving, an A/B/C chimeric polypeptide is obtained that can induce a protective immune response against all three pathogen types (right panel). In Figure 16B, directed evolution is used with substrate nucleic acids from two pathogen strains (A, B), which encode polypeptides that are protective only against the corresponding pathogen. After directed evolution, the resulting chimeric polypeptide can induce an immune response that is effective against not only the two parental pathogen strains, but also against a third strain of pathogen (C).

Figure 17: Possible factors for determining whether a particular polynucleotide encodes an immunogenic polypeptide having a desired property. Shown here are some of the possible factors that can determine whether a particular polynucleotide encodes an immunogenic polypeptide having a desired property, such as enhanced immunogenicity and/or cross-reactivity. Those sequence regions that positively affect a particular property are indicated as plus signs along the antigen gene, while those sequence regions that have a negative effect are shown as minus signs. A pool of related antigen genes are non-stochastically generated using the methods described herein and screened to obtain those evolved nucleic acids that have gained positive sequence regions and lost negative regions. No pre-existing knowledge as to which regions are positive or negative for a particular trait is required.

Figure 18: Screening strategy for antigen library screening. Shown here is a schematic representation of the screening strategy for antigen library screening.

Figure 19: Strategy for pooling and deconvolution as used in antigen library screening. Shown here is a schematic representation of a strategy for pooling and deconvolution as used in antigen library screening.

Figure 20: Exemplary Embodiments of Site-Saturation Mutagenesis.

Figure 21. Schematic representation of a multimodule genetic vaccine vector. Shown here is a schematic representation of a multimodule genetic vaccine vector. A typical genetic vaccine vector will include one or more of the components indicated, each of which can be native or optimized using the directed evolution methods described herein. These directed evolution methods can include the introduction of point mutations by stochastic methods &/or by non-stochastic methods, including "gene site saturation mutagenesis" as described herein. These directed evolution methods can also include stochastic polynucleotide reassembly methods, for example by interrupted synthesis (as described in US5965408). These directed evolution methods can also include non-stochastic polynucleotide reassembly methods as described herein, including synthetic ligation polynucleotide reassembly as described herein. The components can be present on the same vaccine vector, or can be included in a genetic vaccine as separate molecules.

Figure 22A and Figure 22B. Generation of vectors with multiple T cell epitopes. Shown here are two different strategies for generating vectors that contain multiple T cell epitopes obtained, for example, by directed evolution. In Figure 60A, each individual non-

stochastically generated epitope-encoding gene is linked to a single promoter, and multiple promoter-epitope gene constructs can be placed in a single vector. The scheme shown, described &/or referenced herein (including incorporated by reference) involves linking multiple epitope-encoding genes to a single promoter.

Figure 23. Generation of optimized genetic vaccines by directed evolution. Shown here is a diagram of the application of directed evolution to the generation of optimized genetic vaccines. Different forms of polynucleotides having known functional properties (e.g., regulatory, coding, and the like) are evolved and screened to identify variants that exhibit improved properties for use as genetic vaccines.

Figure 24. Recursive application of directed evolution and selection of evolved promoter sequences as an example of flow cytometry-based screening methods. Shown here is a diagram of flow cytometry-based screening methods (FACS) for selection of optimized promoter sequences evolved using recursive applications of the directed evolution methods as described herein. A cytomegalovirus (CMV) promoter is used for illustrative purposes.

Figure 25. An apparatus for microinjections of skin and muscle. Shown here is an apparatus that is suitable for microinjection of genetic vaccines and other reagents into tissue such as skin and muscle. The apparatus is particularly useful for screening large numbers of agents in vivo, being based on a 96-well format. The tips of the apparatus are movable to allow adjustment so that the tips fit into a microtiter plate. After obtaining a reagent of interest is obtained from a plate, the tips are adjusted to a distance of about 2-3 min apart, enabling transfer of 96 different samples to an area of about 1.6 cm by 2.4 cm to about 2.4 cm by 3.6 cm. If desired, the volume of each sample transferred can be electronically controlled; typically the volumes transferred range from about 2 ul to about 5 ul. Each reagent can be mixed with a marker agent or dye to facilitate recognition of the injection site in the tissue. For example, gold particles of different sizes and shaped can be mixed with the reagent of interest, and microscopy and immunohistochemistry used to identify each injection site and to study the reaction induced by each reagent. When muscle tissue is injected, the injection site is first revealed by surgery.

Figure 26. Polynucleotide reassembly. Shown in Panel A is an example of directed evolution. n different strains of a virus are used in this illustration, but the technique is applicable to any single nucleic acid as well as to any nucleic acid for which different strains, species, or gene families have homologous nucleic acids that have one or more nucleotide

changes compared to other homologous nucleic acids. The different variant nucleic acids are experimentally generated, in one aspect, non-stochastically, as described herein, and screened or selected to identify those variants that exhibit the desired property. The directed evolution method(s) and screening can be repeated one or more times to obtain further improvement. Panel B shows that successive rounds of directed evolution can produce progressively enhanced properties, and that the combination of individual beneficial mutations can lead to an enhance improvement compared to the improvement achieved by an individual beneficial mutation.

Figure 27. Vector for promoter evolution. Shown here is an example of a vector that is useful for screening to identify improved promoters from a library of promoter nucleic acids evolved using the directed evolution methods as described herein. Experimentally generated putative promoters are inserted into the vector upstream of a reporter gene for which expression is readily detected. For many applications, it is desirable that the product of the reporter gene be a cell surface protein so that cells which express high levels of the reporter gene can be sorted using flow cytometry-based cell sorting using the reporter gene product. Examples of suitable reporter genes include, for example, B7-2 and mAb179 epitopes. A polyadenylation region is typically placed downstream of the reporter gene (SV40 polyA is illustrated). The vector can also include a second reporter gene an internal control (GFP; green fluorescent protein); this gene is linked to a promoter (SRαp). The vector also typically includes a selectable marker (kanamycin/neomycin resistance is shown), and origins of replication that are functional in mammalian (SV40 ori) and/or bacterial (pUC ori) cells.

Figure 28. Iterative evolution of inducible promoters using directed evolution and flow cytometry-based selection. Shown here is a diagram of a scheme for iterative evolution of inducible promoters using the directed evolution methods as described herein and flow cytometry-based selection. A library of experimentally generated (i.e. produced by one or more directed evolution methods as descried herein) promoter nucleic acids present in appropriate vectors is transfected into the cells, and those cells which exhibit the least expression of marker antigen when grown under uninduced conditions are selected. The vectors (&/or cells containing them) are recovered, and the vectors are introduced into cells (if not contained therein already), and grown under inducing conditions. Those cells that express the highest level of marker antigen are selected.

Figure 29. Evolving a genetic vaccine vector for Oral, Intravenous, Intramuscular, Intradermal, Anal, Vaginal, or Topical Delivery. Illustrated is a strategy for screening of M13 libraries (e.g. generated experimentally using directed evolution as descried herein) for desired targeting of various tissues. The particular example shown here is a schematic diagram of a method for evolving a genetic vaccine vector for improved oral delivery. This may comprise selecting for stability under the acidic conditions of the stomach, and resistance to other degradatory factors of the digestive tract. The particular example illustrated relates to screening for improved oral delivery, but the same principle applies to libraries administered by other routes, including intravenously, intramuscularly, intradermally, anally, vaginally, or topically. After delivery to a test animal, the M13 phage (or a product thereof) is recovered from the tissue of interest. The procedure can be repeated to obtain further optimization.

Figure 30. An alignment of the nucleotide sequences of two human CMV strains and one monkey strain. Shown here is an alignment of the nucleotide sequences of two human cytomegalovirus (CMV) strains and one monkey (Rhesus) strains. This alignment is serviceable for performing non-stochastic polynucleotide reassembly. Nucleotide sequences shared by 2 sequences are in blue lettering & nucleotide sequences shared by 3 sequences are in red lettering to illustrate exemplary but non-limiting examples of reassembly points.

Figure 31. An alignment of IL-4 nucleotide sequences from 3 species (human, primate, and canine). Shown here is an alignment of the IL-4 nucleotide sequences of human, dog and primate strains. This alignment is serviceable for performing non-stochastic polynucleotide reassembly. Nucleotide sequences shared by 2 sequences are in blue lettering & nucleotide sequences shared by 3 sequences are in red lettering to illustrate exemplary but non-limiting examples of reassembly points.

Figure 32. Evolution of polypeptides by synthesizing (in vivo or in vitro) corresponding deduced polynucleotides and subjecting the deduced polynucleotides to directed evolution and expression screening subsequently expressed polypeptides.

Figure 33. Non-stochastic Reassembly of oligo-directed CpG knock-outs. Shown here is a schematic representation of the use of the non-stochastic methods described herein to generate promoter sequences in which unnecessary CpG sequences are deleted, potentially useful CpG sequences are added, and non-replaceable CpG sequences are identified.

Additionally, other sequences (aside from the CpG sequences) can be substituted into, added to, &/or deleted from working polynucleotides.

Figure 34. An Example of a CTIS obtained from HbsAg polypeptide (PreS2 plus S regions). Shown here is an example of a cytotoxic T-cell inducing sequence (CTIS) obtained from HBsAg polypeptide (PreS2 plus S regions).

Figure 35. A CTIS Having Heterologous Epitopes Attached to the Cytoplasmic Portion. Shown here is a CTIS having heterologous epitopes attached to the cytoplasmic portion.

Figure 36. Method for preparing immunogenic agonist sequences (IAS). Shown here is a method for preparing immunogenic agonist sequences (IAS). Wild-type (WT) and mutated forms of nucleic acids encoding a polypeptide of interest are assembled and subjected to non-stochastic reassembly to obtain a nucleic acid encoding a poly-epitope region that contains potential agonist sequences.

Figure 37. Improving Immunostimulatory Sequences (ISS) Using Directed Evolution. Shown here is a scheme for improving immunostimulatory sequences by the directed evolution methods described herein. Oligonucleotide building blocks (e.g. synthetically generated), oligos with known ISS, CpG containing hexamers &/or oligos containing CpG containing hexamers, poly A, C, G, T, etc…can be assembled. The resultant molecule(s) can then by subjected to 1 or more directed evolution methods as described herein.

Figure 38. Screening to identify IL-12 genes that encode recombinant IL-12 having an increased ability to induce T Cell proliferation. Shown here is a diagram of a procedure by which experimentally generated molecules, e.g. non-stochastically generated libraries of human IL-12 genes can be screened to identify evolved IL-12 genes that encode evolved forms of IL-12 having increased ability to induce T cell proliferation.

Figure 39. Model of induction of T cell activation or anergy by genetic vaccine vectors encoding different CD80 and/or CD86 variants. Shown here is a model of how T cell activation or anergy can be induced by genetic vaccine vectors that encode different B7-1 (CD80) and/or B7-2 (CD86) variants.

Figure 40. Screening of CD80/CD86 variants that have improved capacity to induce T cell activation or anergy. Shown here is a method for using directed evolution as described

herein to obtain CD80/CD86 variants that have improved capacity to induce T cell activation or anergy.

Figure 41. An alignment of two CMV-derived nucleotide sequences from human and primate species. Shown here is an alignment of two CMV-derived nucleotide sequences of human and primate strains. This alignment is serviceable for performing non-stochastic polynucleotide reassembly. Nucleotide sequences shared by 2 sequences are in red lettering to illustrate exemplary but non-limiting examples of reassembly points.

Figure 42: An alignment of the IFN-gamma nucleotide sequences from human, cat, rodent species. Shown here is an alignment of the IFN-gamma nucleotide sequences from human, cat, and rodent species. This alignment is serviceable for performing non-stochastic polynucleotide reassembly. Nucleotide sequences shared by 2 sequences are in blue lettering & nucleotide sequences shared by 3 sequences are in red lettering to illustrate exemplary but non-limiting examples of reassembly points.

Figure 43 is a schematic summarizing exemplary applications of the novel capillary array of the invention, e.g., GIGAMATRIX™, Diversa Corporation, San Diego, CA.

Figure 44 is a schematic showing use of paramagnetic beads with the methods of the invention.

Figure 45 is a schematic showing an exemplary use of paramagnetic beads with the methods of the invention. Figure 46 is a schematic summarizing exemplary applications of the novel capillary array of the invention, e.g., GIGAMATRIX™, Diversa Corporation, San Diego, CA. Figure 47 is a schematic summarizing exemplary applications of the novel Gene Site Saturation Mutagenesis method of the invention, as described in detail, below.

Figure 48 is a schematic summarizing exemplary applications of the novel GENE-REASSEMBLY™ method of the invention, as described in detail, below.

Figure 49 is a schematic summarizing exemplary applications of the novel GENE-REASSEMBLY™ method of the invention, as described in detail, below.

Figure 50 is a schematic summarizing an exemplary application of the novel GENE-REASSEMBLY™ method of the invention, as described in detail, below.

Figure 51 is a schematic summarizing an exemplary application of the novel GENE-REASSEMBLY™ method of the invention.

Figure 52 is a schematic summarizing an exemplary application ("dehalogenase reassembly") of the novel GENE-REASSEMBLY™ method of the invention.

Figure 53 is a schematic summarizing the novel TUNEABLE-GENE-REASSEMBLY™ method of the invention, as described, below.

Figure 54 is a schematic summarizing the DNACARPENTER™ reassembly control software that can be used with the methods of the invention.

Figure 55 is a schematic summarizing an exemplary gene family reassembly method of the invention.

Figure 56 is a schematic summarizing an exemplary gene family reassembly method of the invention.

Figure 57 is a schematic summarizing exemplary methods of the invention as described in detail, below.

Figure 58 is a schematic summarizing current deficiencies in antibody generation as discussed in detail, below.

Figure 59 is a schematic summarizing antibodies generated by the methods of the invention, e.g., NATUBODIES™, as described in detail, below.

Figure 60 is a schematic summarizing a bivalent human antibody structure, as discussed in detail, below.

Figure 61 is a schematic summarizing exemplary synthetic human antibodies generated by the methods of the invention, as described in detail, below.

Figure 62 is a schematic summarizing an antibody V-region structure and variability, as discussed in detail, below.

Figure 63 is a schematic summarizing antibody variable region structure, as discussed in detail, below.

Figure 64 is a schematic summarizing exemplary synthetic human antibodies, particularly re-engineered CDR regions, generated by the methods of the invention, as described below.

Figure 65 is a schematic summarizing exemplary synthetic *de novo* antibody libraries generated by the methods of the invention, as described in detail, below.

Figure 66 is a schematic summarizing exemplary methods for generating and screening synthetic human antibodies by the methods of the invention, as described below.

Figure 67 is a schematic summarizing an exemplary method of the invention for screening antibodies, as described below.

Figure 68 is a schematic summarizing an exemplary method for generating antibodies by the methods of the invention, including affinity maturation by a combination of methods of the invention, as described below.

Figure 69 is a schematic summarizing an exemplary application of the novel GENE-REASSEMBLY™ method of the invention.

DETAILED DESCRIPTION

The invention provides methods for generating variant antigen binding sites, antibodies and specific domains or fragments of antibodies (e.g., Fab or Fc domains) by altering template nucleic acid by saturation mutagenesis, an optimized directed evolution system, synthetic ligation reassembly, or a combination thereof. Polypeptides generated by these methods can be analyzed, e.g., screened for a binding activity (e.g., to an antigen), using a novel capillary array platform of the invention.

The invention provides for the modification (e.g., mutagenesis) of Fc domains. In alternative aspects, this invention provides for mutagenizing a percentage, including at least every integer value (i.e. at least 1%, at least 2%, at least 3%, . . . , to at least 99%, or, 100%) of an Fc region or of an Fc-region containing molecule, or fragment, domain or subsection thereof. For example, a nucleic acid encoding an Fc domain (or subsequence thereof) can be modified such that it gains, loses or acquires a modified function (e.g., a binding property) or property (e.g., solubility or antigenicity), or has a modified (e.g., higher or lower) binding affinity. For example, the ability (including, e.g., affinity) of an Fc to bind a particular cell surface receptor (e.g., an Fc receptor, or FcR on, e.g., a B cell, T cell, macrophage, monocyte, mast cell, basophil, dendritic cell, Langerhan cell and the like, or a complement protein or receptor) can be targeted. The Fc domain can be changed to bind a different cell surface receptor (e.g., changed to bind a B cell FcR when the wild type Fc bound to a macrophage FcR or a mast cell FcR), an additional receptor (added function) or fewer receptors (loss of function). The Fc domain can be changed such that it binds to a different complement polypeptide (e.g., changing specificities) or an added specificity or an alterred affinity to a complement protein.

For any protein with an antigen binding site, including antibodies, T cell receptors and Fc domains, their crystal structures can be used to predict which residues may be desirable for targeting. For example, nucleic acid residues encoding solvent exposed

residues, e.g., those involved in protein:protein binding, can be targeting by the methods of the invention. In yet another aspect, the saturation mutagenesis methods of this invention provide mutagenizing (e.g. saturation mutagensis, GSSM) solvent-exposed amino acids of a region (e.g. Fc); e.g., the mutagenesis (e.g. saturation mutagenesis, GSSM) is performed on solvent-exposed amino acids, including those that have been characterized as having a desirable property by, e.g., single codon mutagenesis (e.g. alanine scanning). See, e.g., U.S. Patent No. 5,834,597.

Accordingly, this invention provides a method for making (as well as the product of the method) a library of variants (e.g. generated by saturation mutgenesis) of an antibody comprising a human immunoglobulin (Ig) Fc region (including IgG, IgE, IgA, IgM, IgD). In one aspect, for human IgG, the invention provides variants comprising from at least 2 amino acid substitutions (and every integer value including up to all 19 naturally-occurring amino acid substitutions) at amino acid position 329, or at two or all of amino acid positions 329, 331 and 322 of the human IgG Fc region, where the numbering of the residues in the IgG Fc region is that of the EU index as in Kabat (see also U.S. Patent Nos. 6,242,195 and 6,194,551; and WO 99/51642) and wherein the variant retains the ability to bind antigen. In one aspect of this invention, an antibody comprising a human IgG Fc region is an antibody comprising a human IgG1 Fc region. For example, this invention provides a method for modifying an antibody comprising a human IgG Fc region, the method comprising making (including making and screening) a library of variants having amino acid substitutions at amino acid position (or residue) 329, or at two or all of amino acid positions 329, 331 and 322 of the human IgG Fc region, where the numbering of the residues in the IgG Fc region is that of the EU index as in Kabat, and wherein the variant retains the ability to bind antigen.

This invention also protects the screening of the library of variants, as well as variants identified therefrom. In one aspect, the screening criterion is the selection for a variant that does not activate complement. In another aspect, a screening criterion is the selection for a variant that binds an FcR. In another aspect, a screening criterion is the selection for a variant that binds an FcR, such as FcRI, FcRII, FcRIII or FcRn. In one aspect, a screening criterion is the selection for a variant of an antibody comprising a human IgG Fc region, which variant does not activate complement and comprises an amino acid substitution at amino acid position 322 or amino acid position 329, or both amino acid positions of the human IgG Fc region, where the numbering of the residues in the IgG Fc region is that of the

EU index as in Kabat, and wherein the variant retains the ability to bind antigen. Also provided is a composition of the invention and a physiologically acceptable carrier, for example, a composition of the invention comprising any variant described herein and a physiologically acceptable carrier.

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

The term "nucleic acid" as used herein refers to a deoxyribonucleotide (DNA) or ribonucleotide (RNA) in either single- or double-stranded form. The term encompasses nucleic acids containing known analogues of natural nucleotides. The term encompasses mixed oligonucleotides comprising an RNA portion bearing 2'-O-alkyl substituents conjugated to a DNA portion via a phosphodiester linkage, see, e.g., U.S. Patent No. 5,013,830. The term also encompasses nucleic-acid-like structures with synthetic backbones. DNA backbone analogues provided by the invention include phosphodiester, phosphorothioate, phosphorodithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-N-carbamate, morpholino carbamate, and peptide nucleic acids (PNAs); see Oligonucleotides and Analogues, a Practical Approach, edited by F. Eckstein, IRL Press at Oxford University Press (1991); Antisense Strategies, Annals of the New York Academy of Sciences, Volume 600, Eds. Baserga and Denhardt (NYAS 1992); Milligan (1993) J. Med. Chem. 36:1923-1937; Antisense Research and Applications (1993, CRC Press). PNAs contain non-ionic backbones, such as N-(2-aminoethyl) glycine units. Phosphorothioate linkages are described, e.g., by U.S. Patent Nos. 6,031,092; 6,001,982; 5,684,148; see also, WO 97/03211; WO 96/39154; Mata (1997) Toxicol. Appl. Pharmacol. 144:189-197. Other synthetic backbones encompassed by the term include methyl-phosphonate linkages or alternating methylphosphonate and phosphodiester linkages (see, e.g., U.S. Patent No. 5,962,674; Strauss-Soukup (1997) Biochemistry 36:8692-8698), and benzylphosphonate linkages (see, e.g., U.S. Patent No. 5,532,226; Samstag (1996) Antisense Nucleic Acid Drug Dev 6:153-156). The term nucleic acid is used interchangeably with gene, polynucleotide, DNA, RNA, cDNA, mRNA, oligonucleotide primer, probe and amplification product.

The terms "polypeptide," "protein," and "peptide" include compositions of the invention that also include "analogs," or "conservative variants" and "mimetics" or "peptidomimetics" with structures and activity that substantially correspond to the polypeptide from which the variant was derived.

The term "saturation mutagenesis" includes a method that uses degenerate oligonucleotide primers to introduce point mutations into a polynucleotide, as described in detail, below.

The term "optimized directed evolution system" or "optimized directed evolution" includes a method for reassembling fragments of related nucleic acid sequences, e.g., related genes, and explained in detail, below. This invention provides methods for generating variant antigen binding sites, antibodies and specific domains or fragments of antibodies (e.g., Fab or Fc domains) by mutagenizing a template nucleic acid by an optimized directed evolution system.

The term "synthetic ligation reassembly" or "SLR" includes a method of ligating oligonucleotide fragments in a non-stochastic fashion, and explained in detail, below.

The term "antibody" includes a peptide or polypeptide derived from, modeled after or substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof, capable of specifically binding an antigen or epitope, see, e.g. Fundamental Immunology, Third Edition, W.E. Paul, ed., Raven Press, N.Y. (1993); Wilson (1994) J. Immunol. Methods 175:267-73; Yarmush (1992) J. Biochem. Biophys. Methods 25:85-97. One of skill will appreciate that antibody-encoding nucleic acids and polypeptides may be isolated or synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. The term antibody includes antigen-binding portions, i.e., "antigen binding sites," (e.g., fragments, subsequences, complementarity determining regions (CDRs)) that retain capacity to bind antigen, including (i) a Fab fragment, a monovalent fragment consisting of the VL, VH, CL and CH1 domains; (ii) a $F(ab')_2$ fragment, a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; (iii) an Fd fragment consisting of the VH and CH1 domains; (iv) an Fv fragment consisting of the VL and VH domains of a single arm of an antibody, (v) a dAb fragment (Ward et al., (1989) Nature 341:544-546), which consists of a VH domain; and (vi) an isolated complementarity determining region (CDR). Furthermore, although the two domains of the Fv fragment, VL and VH, are coded for by separate genes, they can be joined, using recombinant methods, by

a synthetic linker that enables them to be made as a single protein chain in which the VL and VH regions pair to form monovalent molecules; also known as single chain Fv (scFv); see e.g., Bird (1988) Science 242:423-426; Huston (1988) Proc. Natl. Acad. Sci. USA 85:5879-5883). Single chain antibodies are also included by reference in the term "antibody." Fragments can be prepared by recombinant techniques or enzymatic or chemical cleavage of intact antibodies. The term also includes multivalent antigen-binding proteins, see, e.g., 6,027,725. The term antibody also includes "chimeric" antibodies either produced by the modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies. Such chimeric antibodies can be "humanized antibodies," i.e., where the epitope binding site is generated from an immunized mammal, such as a mouse, and the structural framework is human. Methods for making chimeric, e.g., "humanized," antibodies are well known in the art, see e.g., U.S. Patent Nos. 5,811,522; 5,789,554; Huse (1989) Science 246:1275; Ward (1989) Nature 341:544; Hoogenboom (1997) Trends Biotechnol. 15:62-70; Katz (1997) Annu. Rev. Biophys. Biomol. Struct. 26:27-45. The term also includes human antibody nucleic acids and polypeptides generated by transgenic non-human animals (e.g., mice) capable of producing human antibodies, as described by, e.g., U.S. Patent Nos. 5,939,598; 5,877,397; 5,874,299; 5,814,318. The term antibody also includes epitope binding polypeptides generated using phage display libraries, and variations thereof, as described by, e.g., U.S. Patent Nos. 5,855,885; 6,027,930. See also, discussion below.

The term "major histocompatibility molecule" or "MHC molecule" as used herein includes all Class I and Class II molecules, including alpha and beta chains of class II molecules and beta-2 microglobulin of Class I chains. Human MHC molecules can also be referred to as "Human Leukocyte Antigens" or HLA. Class II MHC molecules are heterodimers displayed on the cell surface of antigen processing/ presenting cells (APCs), that include, e.g., macrophages, monocytes, activated endothelial cells, human B cells. The methods of the invention include modification of any part or all of these polypeptides to, e.g., modify expression, their association with other molecules, such as antigens or co-stimulatory molecules or T cell receptors, and the like. The structures of, and the isolating, making and using MHC molecules are well known in the art, see, e.g., Fundamental Immunology, Third Edition, Paul (ed) Raven Press, Ltd., New York; and, U.S. Patent Nos. 6,232,445; 6,241,985; 6,245,764; 6,248,564.

The term "T cell receptor" or "TCR" as used herein includes all antigen specific T cell receptor molecules. TCRs are heterodimers (alpha and beta chains, or, gamma and delta chains) displayed on the cell surface of T lymphocytes. The TCR binds to antigenic peptides presented in the binding pocket of an MHC molecule. The methods of the invention include modification of any part or all of these polypeptides to, e.g., modify their expression, their association with other molecules, such as antigenic peptides or co-stimulatory molecules or an MHC molecule, and the like. The structures of, and the isolating, making and using MHC molecules are well known in the art, see, e.g., Fundamental Immunology, Third Edition, Paul (ed) Raven Press, Ltd., New York, U.S. Patent Nos. 5,316,925; 5,601,822; 5,614,192; 5,635,363; 5,667,967; 5,840,304; 6,054,292; 6,180,104; 6,245,764.

The invention provides arrays comprising samples of (i.e., subsets of), or all of, the antigen binding sites that are isolated from and/or expressed by an individual, or, complementary to antigen binding sites isolated from and/or expressed by the individual. The invention also provides arrays comprising nucleic acids encoding these antigen binding sites. The present invention can be practiced with any known "array," also referred to as a "microarray" or "DNA array" or "nucleic acid array" or "polypeptide array" or "biochip," or variation thereof. In practicing the methods of the invention, known arrays and methods of making and using arrays can be incorporated in whole or in part, or variations thereof, as described, for example, in U.S. Patent Nos. 6,277,628; 6,277,489; 6,261,776; 6,258,606; 6,054,270; 6,048,695; 6,045,996; 6,022,963; 6,013,440; 5,965,452; 5,959,098; 5,856,174; 5,830,645; 5,770,456; 5,632,957; 5,556,752; 5,143,854; 5,807,522; 5,800,992; 5,744,305; 5,700,637; 5,556,752; 5,434,049; see also, e.g., WO 99/51773; WO 99/09217; WO 97/46313; WO 96/17958; see also, e.g., Johnston (1998) Curr. Biol. 8:R171-R174; Schummer (1997) Biotechniques 23:1087-1092; Kern (1997) Biotechniques 23:120-124; Solinas-Toldo (1997) Genes, Chromosomes & Cancer 20:399-407; Bowtell (1999) Nature Genetics Supp. 21:25-32. See also published U.S. patent applications Nos. 20010018642; 20010019827; 20010016322; 20010014449; 20010014448; 20010012537; 20010008765.

The term "agent" is used herein to denote a chemical compound, a mixture of chemical compounds, an array of spatially localized compounds (e.g., a VLSIPS peptide array, polynucleotide array, and/or combinatorial small molecule array), biological macromolecule, a bacteriophage peptide display library, a bacteriophage antibody (e.g., scFv) display library, a polysome peptide display library, or an extract made form biological

materials such as bacteria, plants, fungi, or animal (particular mammalian) cells or tissues. Agents are evaluated for potential activity as anti-neoplastics, anti-inflammatories or apoptosis modulators by inclusion in screening assays described hereinbelow. Agents are evaluated for potential activity as specific protein interaction inhibitors (i.e., an agent which selectively inhibits a binding interaction between two predetermined polypeptides but which doe snot substantially interfere with cell viability) by inclusion in screening assays described hereinbelow.

An "ambiguous base requirement" in a restriction site refers to a nucleotide base requirement that is not specified to the fullest extent, i.e. that is not a specific base (such as, in a non-limiting exemplification, a specific base selected from A, C, G, and T), but rather may be any one of at least two or more bases. Commonly accepted abbreviations that are used in the art as well as herein to represent ambiguity in bases include the following: **R** = G or A; **Y** = C or T; **M** = A or C; **K** = G or T; **S** = G or C; **W** = A or T; **H** = A or C or T; **B** = G or T or C; **V** = G or C or A; **D** = G or A or T; **N** = A or C or G or T.

"Alignment" with respect to molecular sequences is a way to determine similarity between 2 or more sequences. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally Ausubel et al., infra).

One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., J Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word

hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915). In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993) Proc. Nat'l. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0. 1, or less than about 0.0 1, or less than about 0.001. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to", refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target polynucleotide sequence.

The term "amino acid" as used herein refers to any organic compound that contains an amino group ($-NH_2$) and a carboxyl group ($-COOH$); in one aspect, either as free groups

or alternatively after condensation as part of peptide bonds. The "twenty naturally encoded polypeptide-forming alpha-amino acids" are understood in the art and refer to: alanine (ala or A), arginine (arg or R), asparagine (asn or N), aspartic acid (asp or D), cysteine (cys or C), gluatamic acid (glu or E), glutamine (gln or Q), glycine (gly or G), histidine (his or H), isoleucine (ile or I), leucine (leu or L), lysine (lys or K), methionine (met or M), phenylalanine (phe or F), proline (pro or P), serine (ser or S), threonine (thr or T), tryptophan (trp or W), tyrosine (tyr or Y), and valine (val or V).

The term "amplification" means that the number of copies of a polynucleotide is increased.

The term "antibody", as used herein, refers to intact immunoglobulin molecules, as well as fragments of immunoglobulin molecules, such as Fab, Fab', $(Fab')_2$, Fv, and SCA fragments, that are capable of binding to an epitope of an antigen. These antibody fragments, which retain some ability to selectively bind to an antigen (*e.g.*, a polypeptide antigen) of the antibody from which they are derived, can be made using well known methods in the art (see, e.g., Harlow and Lane, *supra*), and are described further, as follows.

(1) An Fab fragment consists of a monovalent antigen-binding fragment of an antibody molecule, and can be produced by digestion of a whole antibody molecule with the enzyme papain, to yield a fragment consisting of an intact light chain and a portion of a heavy chain.

(2) An Fab' fragment of an antibody molecule can be obtained by treating a whole antibody molecule with pepsin, followed by reduction, to yield a molecule consisting of an intact light chain and a portion of a heavy chain. Two Fab' fragments are obtained per antibody molecule treated in this manner.

(3) An $(Fab')_2$ fragment of an antibody can be obtained by treating a whole antibody molecule with the enzyme pepsin, without subsequent reduction. A $(Fab')_2$ fragment is a dimer of two Fab' fragments, held together by two disulfide bonds.

(4) An Fv fragment is defined as a genetically engineered fragment containing the variable region of a light chain and the variable region of a heavy chain expressed as two chains.

(5) An single chain antibody ("SCA") is a genetically engineered single chain molecule containing the variable region of a light chain and the variable region of a heavy chain, linked by a suitable, flexible polypeptide linker.

The term “Applied Molecular Evolution” ("AME") means the application of an evolutionary design algorithm to a specific, useful goal. While many different library formats for AME have been reported for polynucleotides, peptides and proteins (phage, lacI and polysomes), none of these formats have provided for recombination by random cross-overs to deliberately create a combinatorial library.

A molecule that has a “chimeric property” is a molecule that is: 1) in part homologous and in part heterologous to a first reference molecule; while 2) at the same time being in part homologous and in part heterologous to a second reference molecule; without 3) precluding the possibility of being at the same time in part homologous and in part heterologous to still one or more additional reference molecules. In a non-limiting embodiment, a chimeric molecule may be prepared by assemblying a reassortment of partial molecular sequences. In a non-limiting aspect, a chimeric polynucleotide molecule may be prepared by synthesizing the chimeric polynucleotide using plurality of molecular templates, such that the resultant chimeric polynucleotide has properties of a plurality of templates.

The term "cognate" as used herein refers to a gene sequence that is evolutionarily and functionally related between species. For example, but not limitation, in the human genome the human CD4 gene is the cognate gene to the mouse 3d4 gene, since the sequences and structures of these two genes indicate that they are highly homologous and both genes encode a protein which functions in signaling T cell activation through MHC class II-restricted antigen recognition.

A "comparison window," as used herein, refers to a conceptual segment of at least 20 contiguous nucleotide positions wherein a polynucleotide sequence may be compared to a reference sequence of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith (Smith and Waterman, *Adv Appl Math*, 1981; Smith and Waterman, J Teor Biol, 1981; Smith and Waterman, *J Mol Biol*, 1981; Smith et al, *J Mol Evol*, 1981), by the homology alignment algorithm of Needleman (Needleman and Wuncsch, 1970), by the search of similarity method of Pearson (Pearson and Lipman, 1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA

in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

As used herein, the term "complementarity-determining region" and "CDR" refer to the art-recognized term as exemplified by the Kabat and Chothia CDR definitions also generally known as supervariable regions or hypervariable loops (Chothia and Lesk, 1987; Clothia et al, 1989; Kabat et al, 1987; and Tramontano et al, 1990). Variable region domains typically comprise the amino-terminal approximately 105-115 amino acids of a naturally-occurring immunoglobulin chain (e.g., amino acids 1-110), although variable domains somewhat shorter or longer are also suitable for forming single-chain antibodies.

"Conservative amino acid substitutions" refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Exemplary conservative amino acids substitution groups are : valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

"Conservatively modified variations" of a particular polynucleotide sequence refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of "conservatively modified variations." Every polynucleotide sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill will

recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence. Furthermore, one of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following five groups each contain amino acids that are conservative substitutions for one another: Aliphatic: Glycine (G), Alanine (A), Valine (V), Leucine (L), Isoleucine (1); Aromatic: Phenylalanine (F), Tyrosine (Y), Tryptophan (W); Sulfur-containing: Methionine (M), Cysteine (C); Basic: Arginine (R), Lysine (K), Histidine (H); Acidic: Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q). See also, Creighton (1984) Proteins, W.H. Freeman and Company, for additional groupings of amino acids. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations".

The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (i.e., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference "TATAC" and is complementary to a reference sequence "GTATA."

The term "cytokine" includes, for example, interleukins, interferons, chemokines, hematopoietic growth factors, tumor necrosis factors and transforming growth factors. In general these are small molecular weight proteins that regulate maturation, activation, proliferation and differentiation of the cells of the immune system.

The term "degrading effective" amount refers to the amount of enzyme which is required to process at least 50% of the substrate, as compared to substrate not contacted with the enzyme. In one aspect, at least 80% of the substrate is degraded.

As used herein, the term "defined sequence framework" refers to a set of defined sequences that are selected on a non-random basis, generally on the basis of experimental data or structural data; for example, a defined sequence framework may comprise a set of amino acid sequences that are predicted to form a ß-sheet structure or may comprise a leucine zipper heptad repeat motif, a zinc-finger domain, among other variations. A "defined sequence kernal" is a set of sequences which encompass a limited scope of variability. Whereas (1) a completely random 10-mer sequence of the 20 conventional amino acids can be any of $(20)^{10}$ sequences, and (2) a pseudorandom 10-mer sequence of the 20 conventional amino acids can be any of $(20)^{10}$ sequences but will exhibit a bias for certain residues at certain positions and/or overall, (3) a defined sequence kernal is a subset of sequences if each residue position was allowed to be any of the allowable 20 conventional amino acids (and/or allowable unconventional amino/imino acids). A defined sequence kernal generally comprises variant and invariant residue positions and/or comprises variant residue positions which can comprise a residue selected from a defined subset of amino acid residues), and the like, either segmentally or over the entire length of the individual selected library member sequence. Defined sequence kernels can refer to either amino acid sequences or polynucleotide sequences. Of illustration and not limitation, the sequences $(NNK)_{10}$ and $(NNM)_{10}$, wherein N represents A, T, G, or C; K represents G or T; and M represents A or C, are defined sequence kernels.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 μg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a gel to isolate the desired fragment.

"Directional ligation" refers to a ligation in which a 5' end and a 3' end of a polynuclotide are different enough to specify an exemplary ligation orientation. For

example, an otherwise untreated and undigested PCR product that has two blunt ends will typically not have an exemplary ligation orientation when ligated into a cloning vector digested to produce blunt ends in its multiple cloning site; thus, directional ligation will typically not be displayed under these circumstances. In contrast, directional ligation will typically displayed when a digested PCR product having a 5' *Eco*R I-treated end and a 3' *Bam*H I-is ligated into a cloning vector that has a multiple cloning site digested with *Eco*R I and *Bam*H I.

The term "DNA shuffling" is used herein to indicate recombination between substantially homologous but non-identical sequences, in some embodiments DNA shuffling may involve crossover via non-homologous recombination, such as via cer/lox and/or flp/frt systems and the like.

As used in this invention, the term "epitope" refers to an antigenic determinant on an antigen, such as a phytase polypeptide, to which the paratope of an antibody, such as an phytase-specific antibody, binds. Antigenic determinants usually consist of chemically active surface groupings of molecules, such as amino acids or sugar side chains, and can have specific three-dimensional structural characteristics, as well as specific charge characteristics. As used herein "epitope" refers to that portion of an antigen or other macromolecule capable of forming a binding interaction that interacts with the variable region binding body of an antibody. Typically, such binding interaction is manifested as an intermolecular contact with one or more amino acid residues of a CDR.

An "exogenous DNA segment", "heterologous sequence" or a "heterologous nucleic acid", as used herein, is one that originates from a source foreign to the particular host cell, or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell, but has been modified. Modification of a heterologous sequence in the applications described herein typically occurs through the use of stochastic (e.g. polynucleotide shuffling & interrupted synthesis) and non-stochastic polynucleotide reassembly. Thus, the terms refer to a DNA segment which is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found.

"Exogenous" DNA segments are expressed to yield exogenous polypeptides.

The term "gene" is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences

required for their expression. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

An "experimentally generated (in vitro &/or in vivo) polynucleotide" (which term includes a "recombinant polynucleotide") or an "experimentally (in vitro &/or in vivo) generated polypeptide" (which term includes a "experimentally generated polypeptide") is a non-naturally occurring polynucleotide or polypeptide that includes nucleic acid or amino acid sequences, respectively, from more than one source nucleic acid or polypeptide, which source nucleic acid or polypeptide can be a naturally occurring nucleic acid or polypeptide, or can itself have been subjected to mutagenesis or other type of modification. The source polynucleotides or polypeptides from which the different nucleic acid or amino acid sequences are derived are sometimes homologous (i.e., have, or encode a polypeptide that encodes, the same or a similar structure and/or function), and are often from different isolates, serotypes, strains, species, of organism or from different disease states, for example.

The terms "fragment", "derivative" and "analog" when referring to a reference polypeptide comprise a polypeptide which retains at least one biological function or activity that is at least essentially same as that of the reference polypeptide. Furthermore, the terms "fragment", "derivative" or "analog" are exemplified by a "pro-form" molecule, such as a low activity proprotein that can be modified by cleavage to produce a mature enzyme with significantly higher activity.

A method is provided herein for producing from a template polypeptide a set of progeny polypeptides in which a "full range of single amino acid substitutions" is represented at each amino acid position. As used herein, "full range of single amino acid substitutions" is in reference to the naturally encoded 20 naturally encoded polypeptide-forming alpha-amino acids, as described herein.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

"Genetic instability", as used herein, refers to the natural tendency of highly repetitive sequences to be lost through a process of reductive events generally involving sequence

simplification through the loss of repeated sequences. Deletions tend to involve the loss of one copy of a repeat and everything between the repeats.

The term "heterologous" means that one single-stranded nucleic acid sequence is unable to hybridize to another single-stranded nucleic acid sequence or its complement. Thus areas of heterology means that areas of polynucleotides or polynucleotides have areas or regions within their sequence which are unable to hybridize to another nucleic acid or polynucleotide. Such regions or areas are for example areas of mutations.

The term "homologous" or "homeologous" means that one single-stranded nucleic acid nucleic acid sequence may hybridize to a complementary single-stranded nucleic acid sequence. The degree of hybridization may depend on a number of factors including the amount of identity between the sequences and the hybridization conditions such as temperature and salt concentrations as discussed later. In one aspect the region of identity is greater than about 5 bp, or, the region of identity is greater than 10 bp.

An immunoglobulin light or heavy chain variable region consists of a "framework" region interrupted by three hypervariable regions, also called CDR's. The extent of the framework region and CDR's have been precisely defined; see "Sequences of Proteins of Immunological Interest" (Kabat et al, 1987). The sequences of the framework regions of different light or heavy chains are relatively conserved within a specie. As used herein, a "human framework region" is a framework region that is substantially identical (about 85 or more, usually 90-95 or more) to the framework region of a naturally occurring human immunoglobulin. the framework region of an antibody, that is the combined framework regions of the constituent light and heavy chains, serves to position and align the CDR's. The CDR's are primarily responsible for binding to an epitope of an antigen.

The benefits of this invention extend to "commercial applications" (or commercial processes), which term is used to include applications in commercial industry proper (or simply industry) as well as non-commercial commercial applications (e.g. biomedical research at a non-profit institution). Relevant applications include those in areas of diagnosis, medicine, agriculture, manufacturing, and academia.

The term "identical" or "identity" means that two nucleic acid sequences have the same sequence or a complementary sequence. Thus, "areas of identity" means that regions or areas of a polynucleotide or the overall polynucleotide are identical or complementary to areas of another polynucleotide or the polynucleotide.

The terms "identical" or percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

A further indication that two nucleic acid sequences or polypeptides are substantially "identical" is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with, or specifically binds to, the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The term "isolated", when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is free of at least one other cellular components with which it is associated in the natural state. It can be substantially is free of at least one other cellular components with which it is associated in the natural state. It can be in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated gene is separated

from open reading frames which flank the gene and encode a protein other than the gene of interest.

By "isolated nucleic acid" is meant a nucleic acid, e.g., a DNA or RNA molecule, that is not immediately contiguous with the 5' and 3' flanking sequences with which it normally is immediately contiguous when present in the naturally occurring genome of the organism from which it is derived. The term thus describes, for example, a nucleic acid that is incorporated into a vector, such as a plasmid or viral vector; a nucleic acid that is incorporated into the genome of a heterologous cell (or the genome of a homologous cell, but at a site different from that at which it naturally occurs); and a nucleic acid that exists as a separate molecule, e.g., a DNA fragment produced by PCR amplification or restriction enzyme digestion, or an RNA molecule produced by *in vitro* transcription. The term also describes a recombinant nucleic acid that forms part of a hybrid gene encoding additional polypeptide sequences that can be used, for example, in the production of a fusion protein.

As used herein "ligand" refers to a molecule, such as a random peptide or variable segment sequence, that is recognized by a particular receptor. As one of skill in the art will recognize, a molecule (or macromolecular complex) can be both a receptor and a ligand. In general, the binding partner having a smaller molecular weight is referred to as the ligand and the binding partner having a greater molecular weight is referred to as a receptor.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Sambrook et al, 1982, p. 146; Sambrook, 1989). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

As used herein, "linker" or "spacer" refers to a molecule or group of molecules that connects two molecules, such as a DNA binding protein and a random peptide, and serves to place the two molecules in an exemplary configuration, e.g., so that the random peptide can bind to a receptor with minimal steric hindrance from the DNA binding protein.

As used herein, a "molecular property to be evolved" includes reference to molecules comprised of a polynucleotide sequence, molecules comprised of a polypeptide sequence, and molecules comprised in part of a polynucleotide sequence and in part of a polypeptide sequence. Particularly relevant - but by no means limiting - examples of molecular

properties to be evolved include enzymatic activities at specified conditions, such as related to temperature; salinity; pressure; pH; and concentration of glycerol, DMSO, detergent, &/or any other molecular species with which contact is made in a reaction environment. Additional particularly relevant - but by no means limiting - examples of molecular properties to be evolved include stabilities - e.g. the amount of a residual molecular property that is present after a specified exposure time to a specified environment, such as may be encountered during storage.

A "multivalent antigenic polypeptide" or a "recombinant multivalent antigenic polypeptide" is a non-naturally occurring polypeptide that includes amino acid sequences from more than one source polypeptide, which source polypeptide is typically a naturally occurring polypeptide. At least some of the regions of different amino acid sequences constitute epitopes that are recognized by antibodies found in a mammal that has been injected with the source polypeptide. The source polypeptides from which the different epitopes are derived are usually homologous (i.e., have the same or a similar structure and/or function), and are often from different isolates, serotypes, strains, species, of organism or from different disease states, for example.

The term "mutations" includes changes in the sequence of a wild-type or parental nucleic acid sequence or changes in the sequence of a peptide. Such mutations may be point mutations such as transitions or transversions. The mutations may be deletions, insertions or duplications. A mutation can also be a "chimerization", which is exemplified in a progeny molecule that is generated to contain part or all of a sequence of one parental molecule as well as part or all of a sequence of at least one other parental molecule. This invention provides for both chimeric polynucleotides and chimeric polypeptides.

As used herein, the degenerate "N,N,G/T" nucleotide sequence represents 32 possible triplets, where "N" can be A, C, G or T.

The term "naturally-occurring" as used herein as applied to the object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses bacteria, protozoa, insects, plants or mammalian tissue) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally occurring. Generally, the term naturally occurring refers to an object as present in a non-pathological (un-diseased) individual, such as would be typical for the species.